Universidad Popular del Cesar

Minería de datos y Patrones

Version 2024-I

Feature Selection
Exhaustive Search

[Capítulo 3]

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

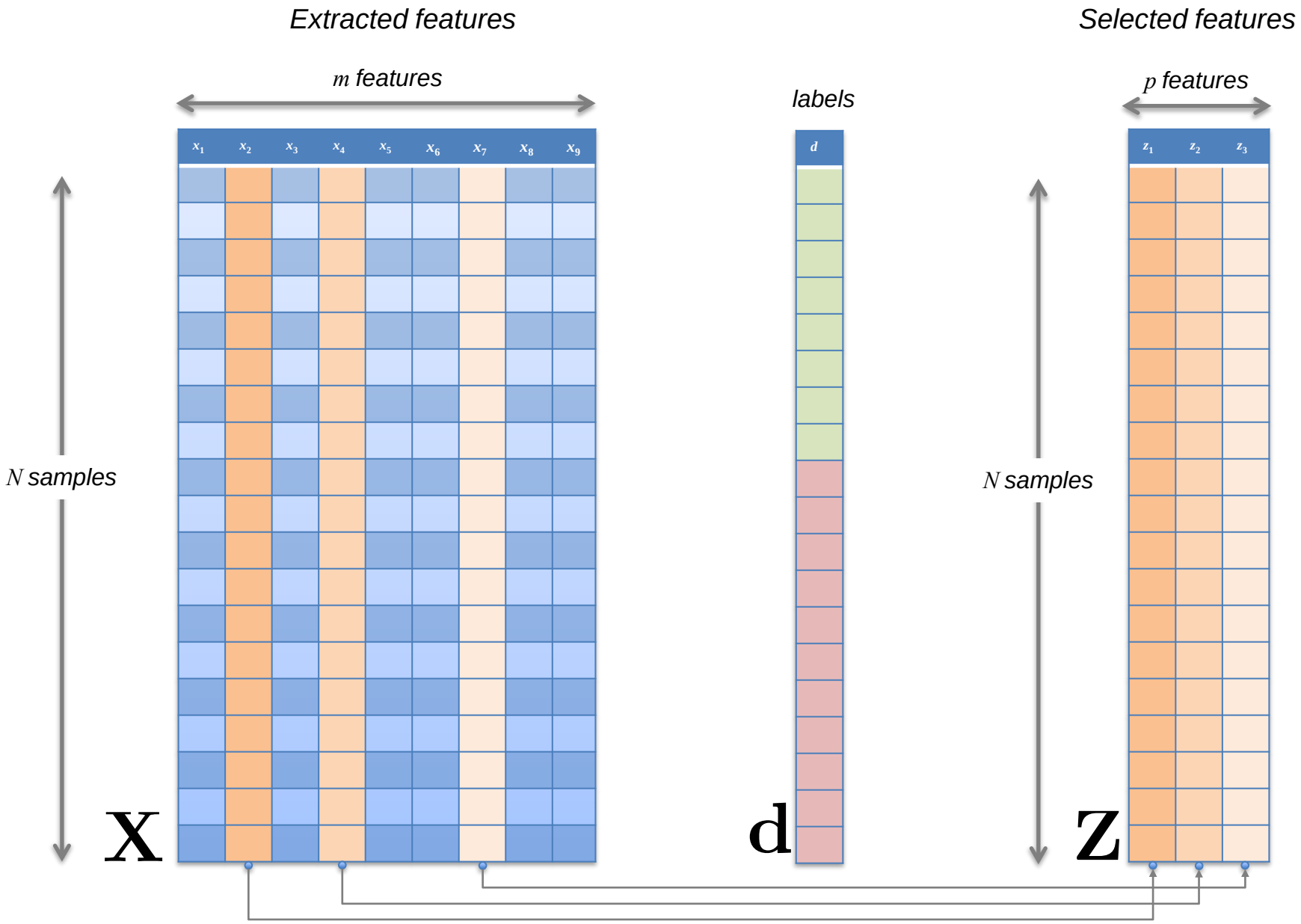
Director Semillero TRIAC

Ingenieria Electronica

Universidad Popular del Cesar

Example:

From 9 features, select 3

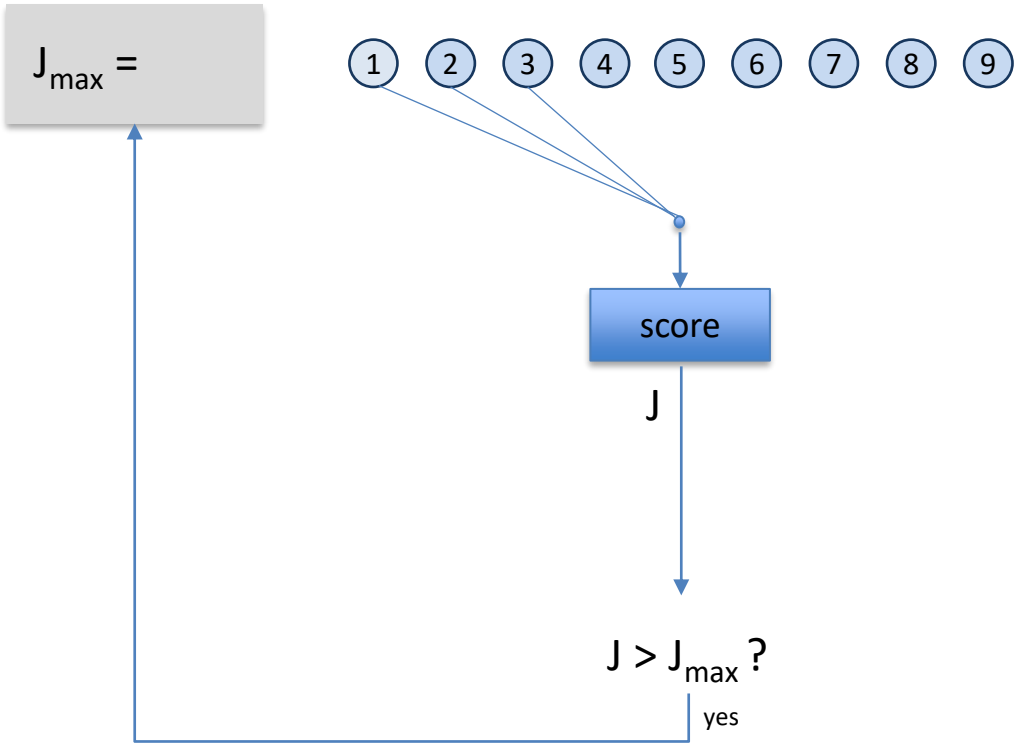

Extracted features
Selected features
m features
p features
labels
N samples
N samples
x1 x2 x3 x4 x5 x6 x7 x8 x9
d
z1 z2 z3
X
d
Z

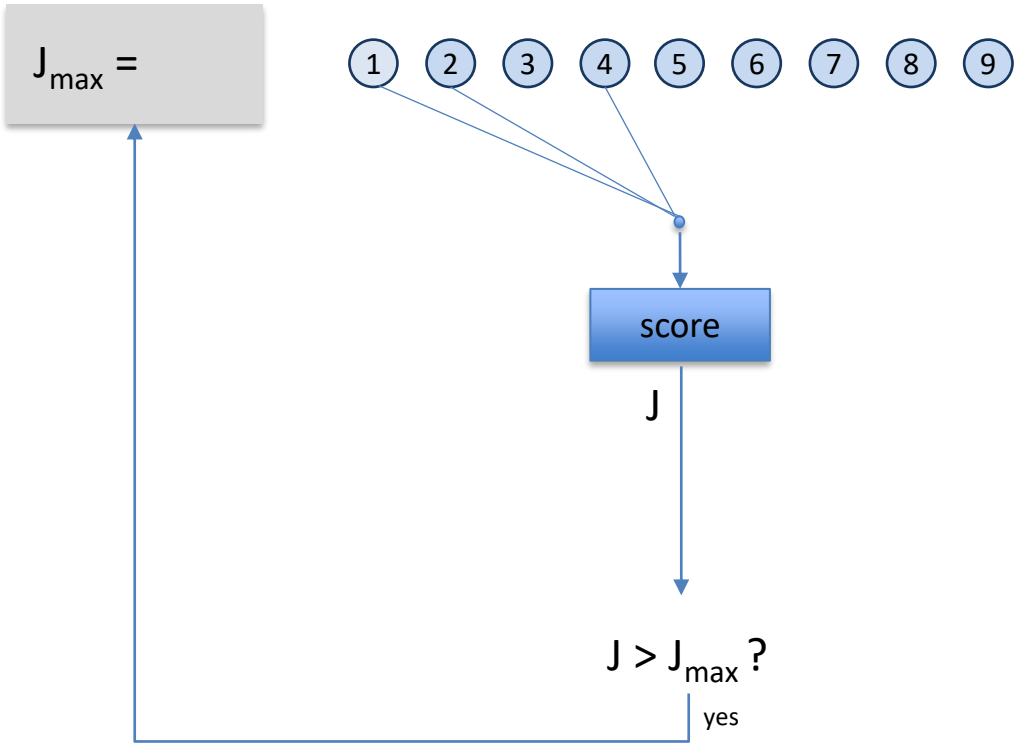
All combinations of three:

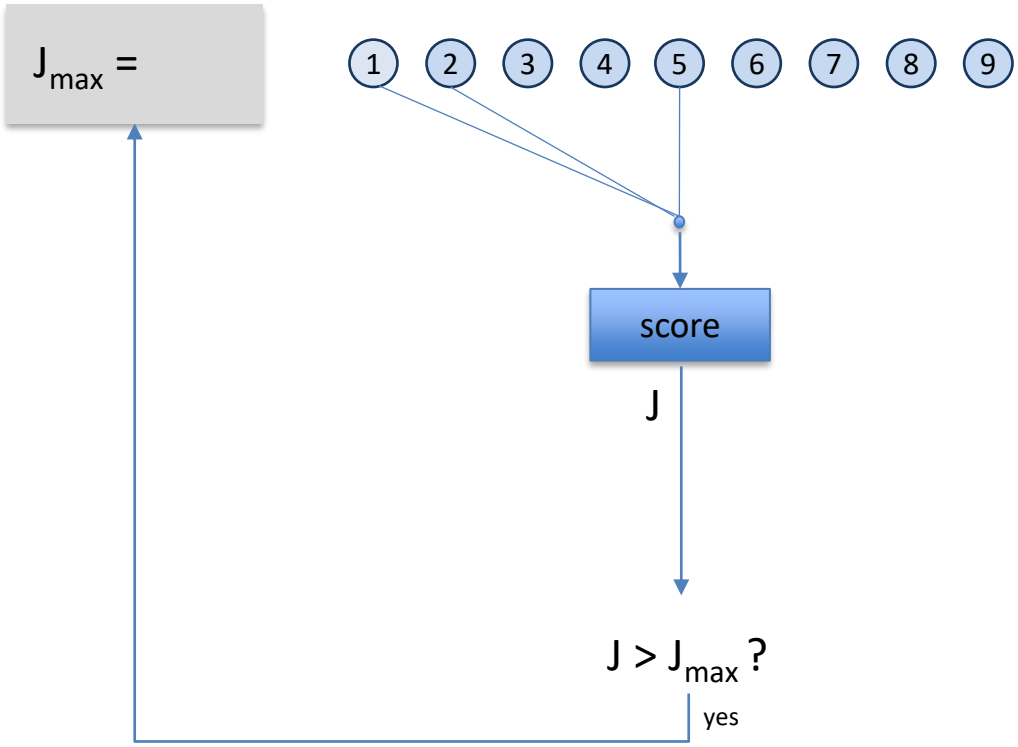
$$\binom{9}{3} = \frac{9!}{6!3!} = 84$$

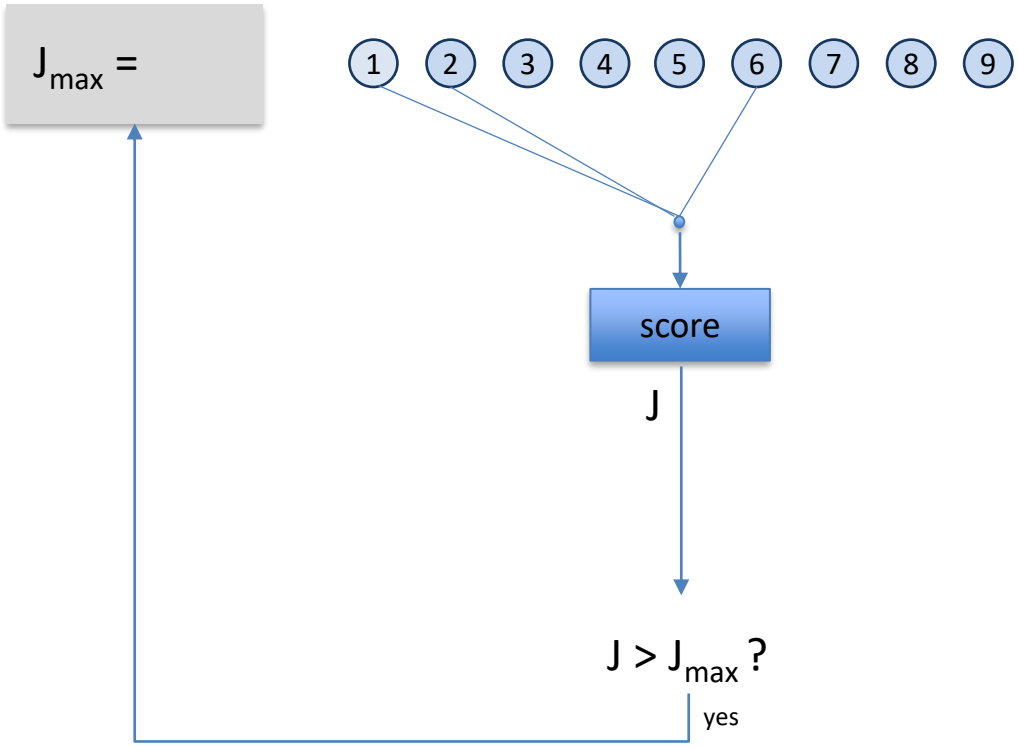
For m = 20, there are 1140 combinations

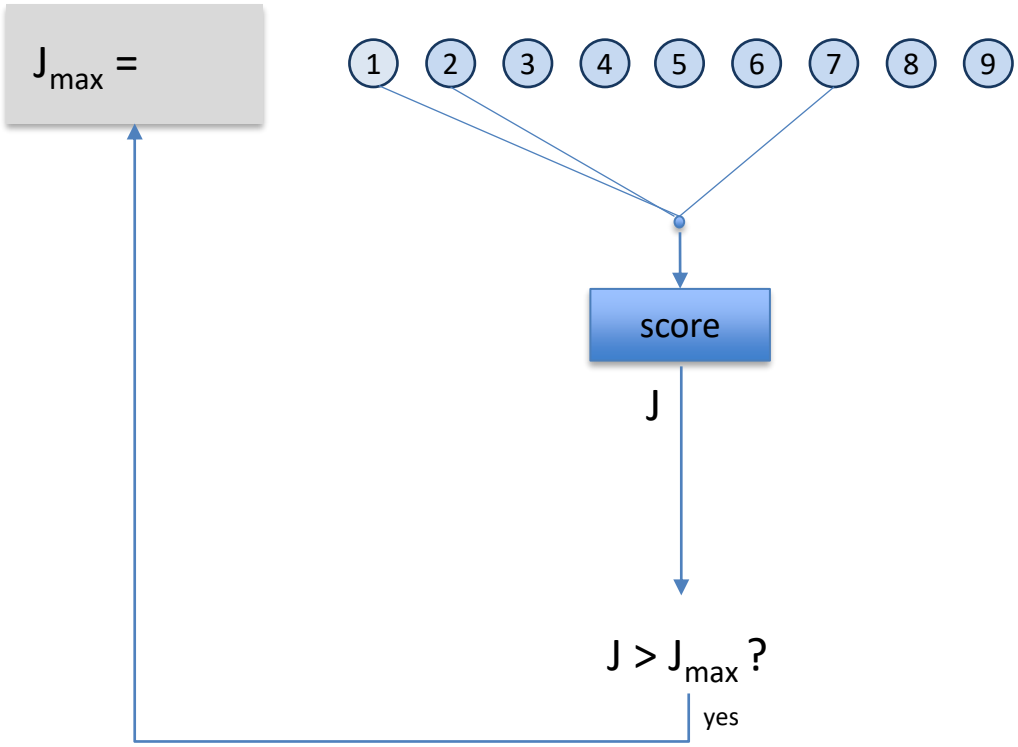
1 2 3
1 2 4
1 2 5
1 2 6
1 2 7
1 2 8
1 2 9
1 3 4
1 3 5
1 3 6
1 3 7
1 3 8
1 3 9
1 4 5
1 4 6
1 4 7
1 4 8
1 4 9
1 5 6
1 5 7
1 5 8
1 5 9
1 6 7

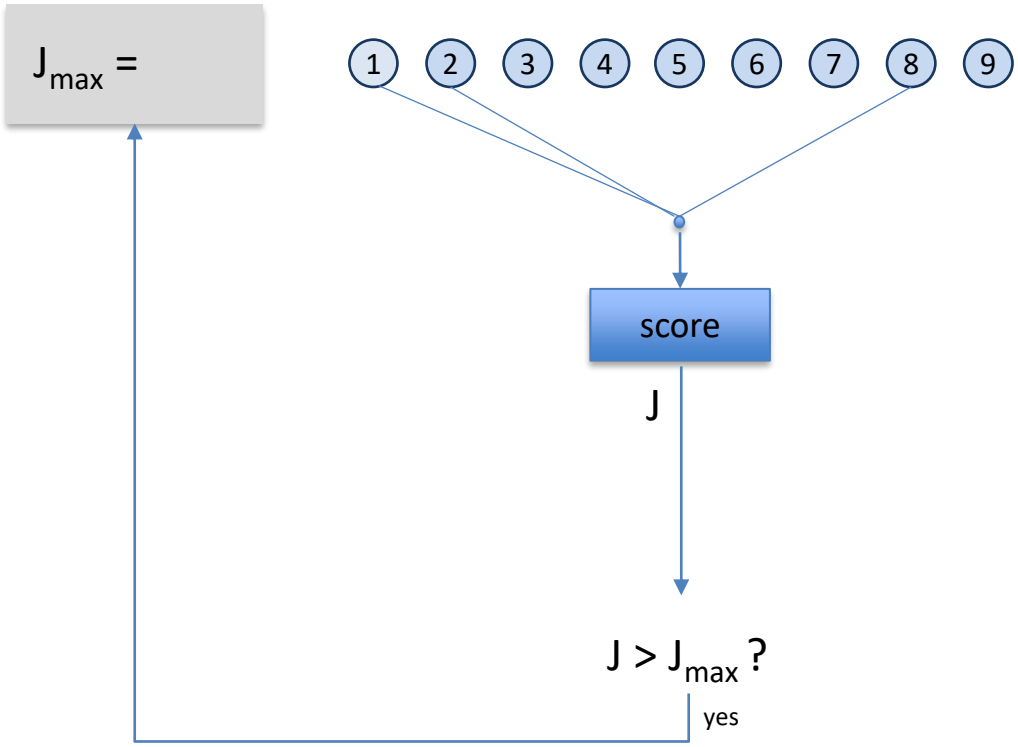

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

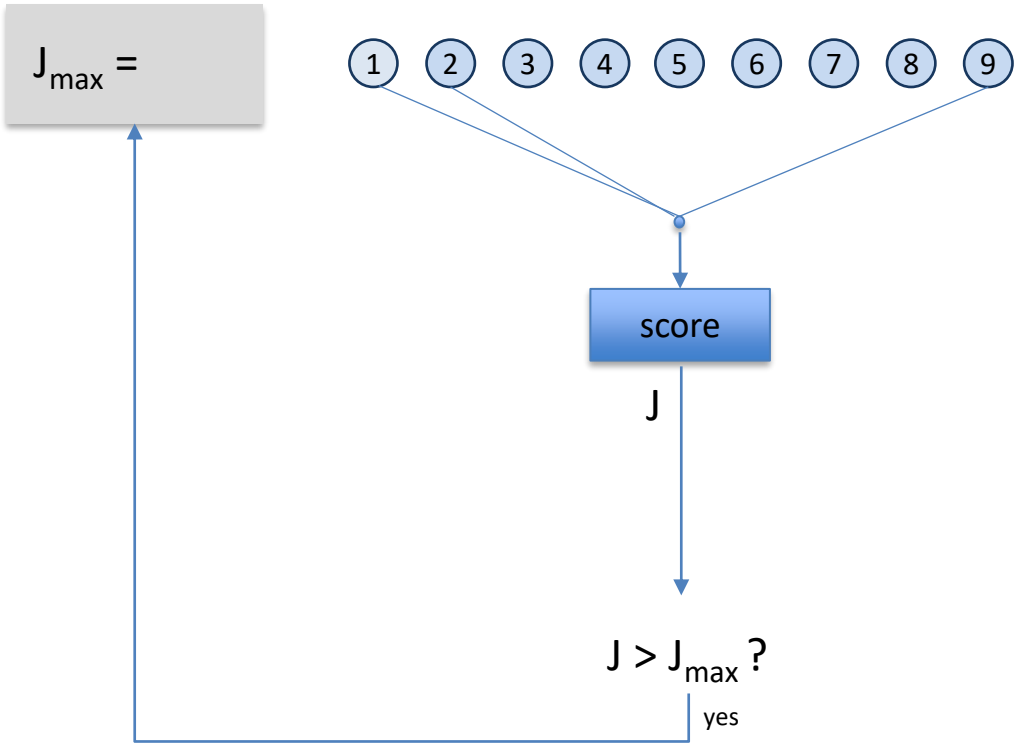

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

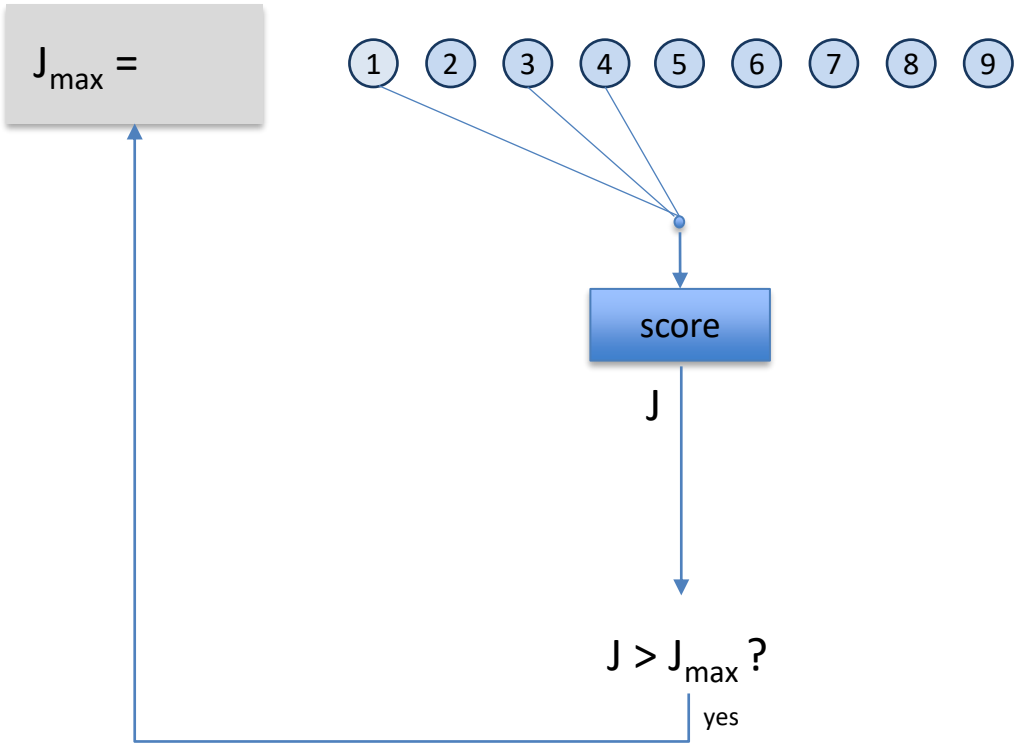

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

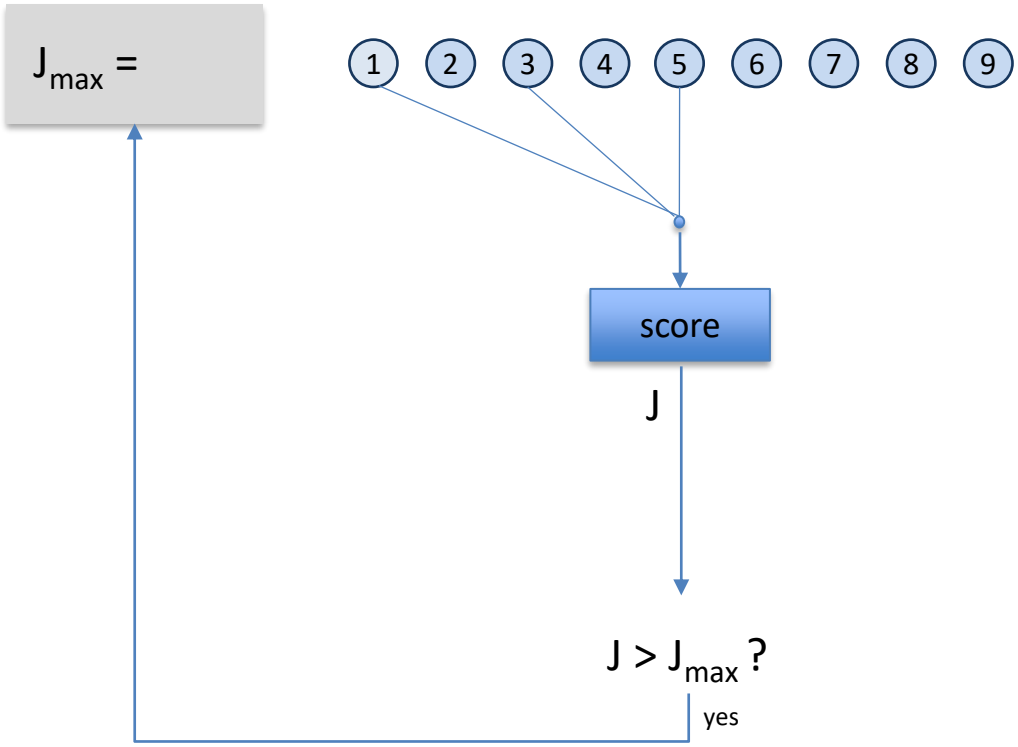

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

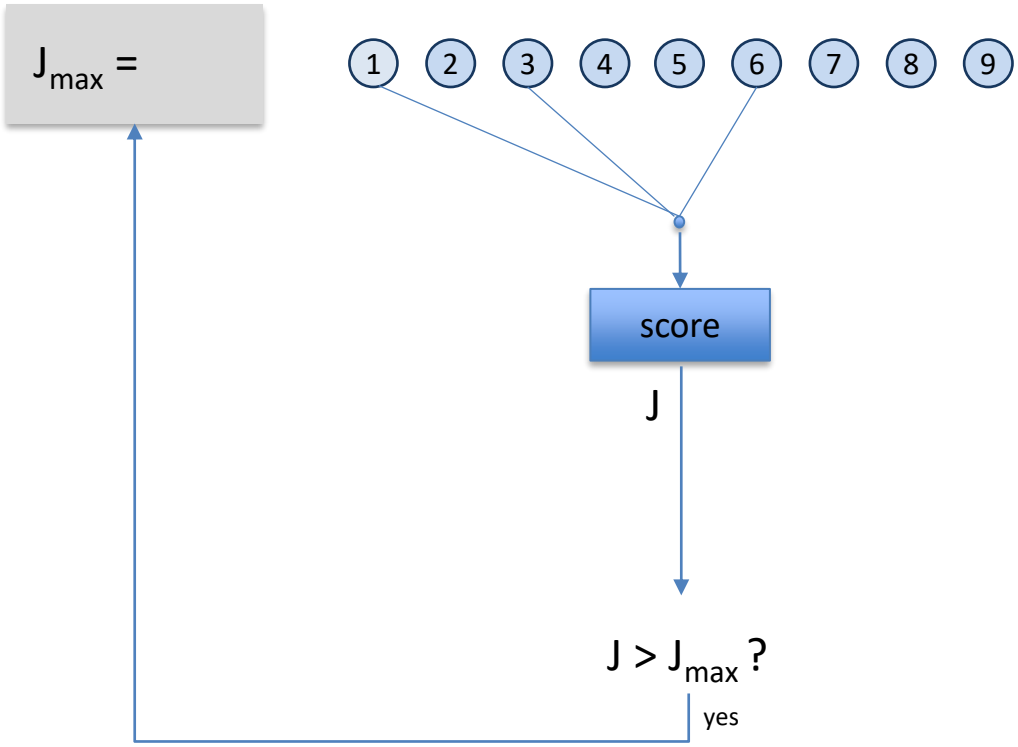

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

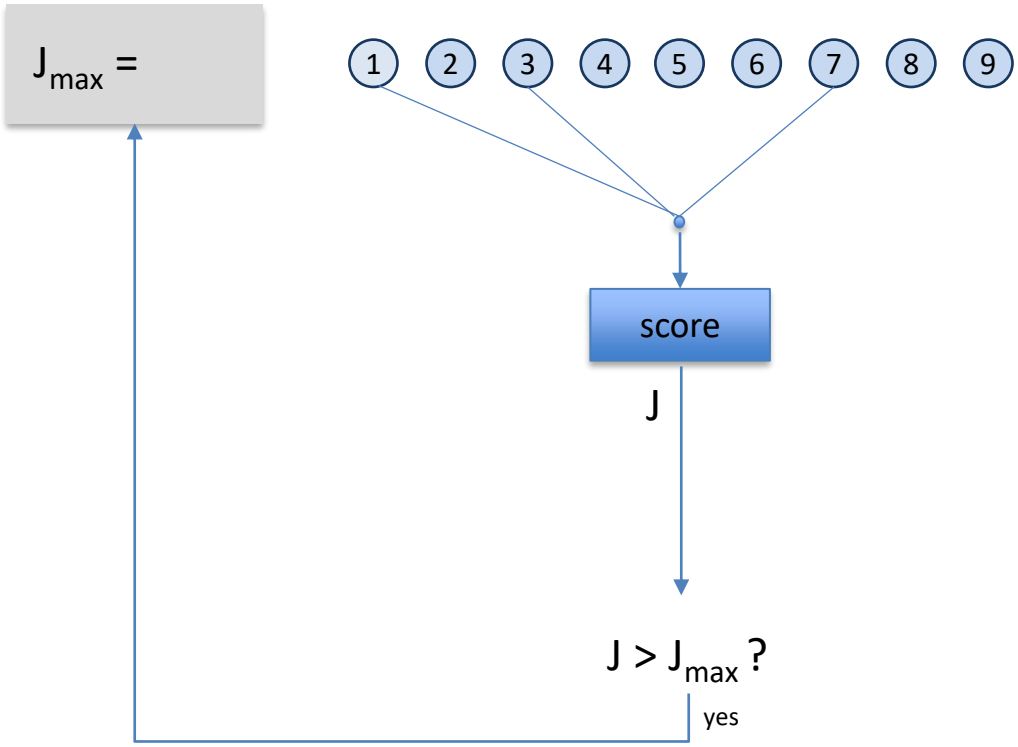

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

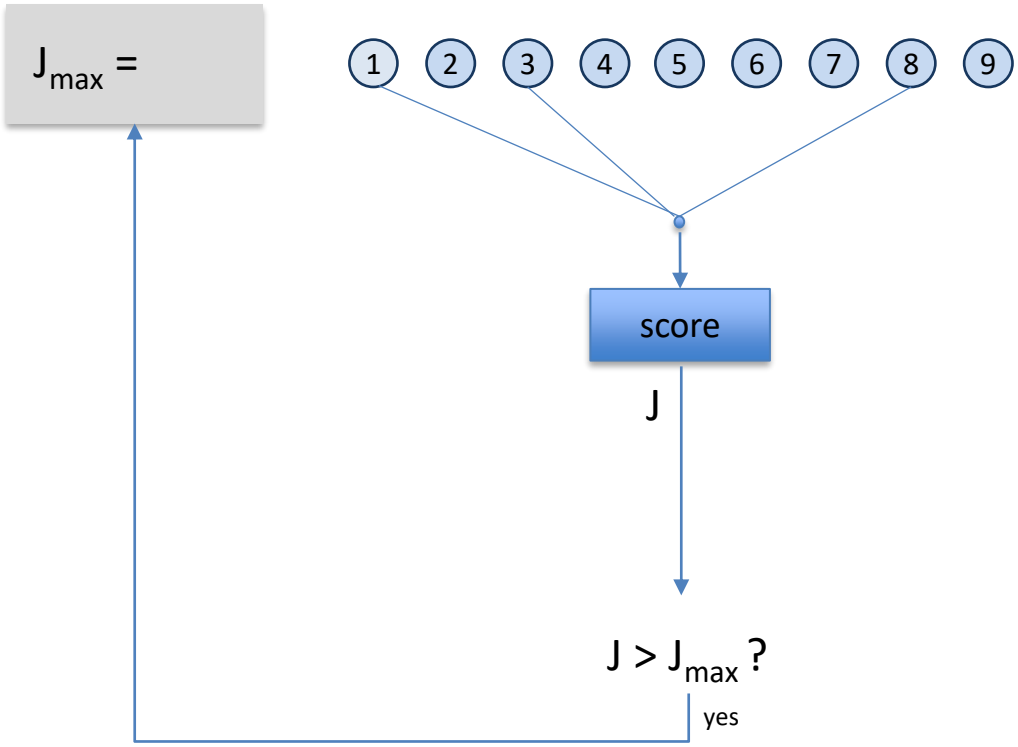

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

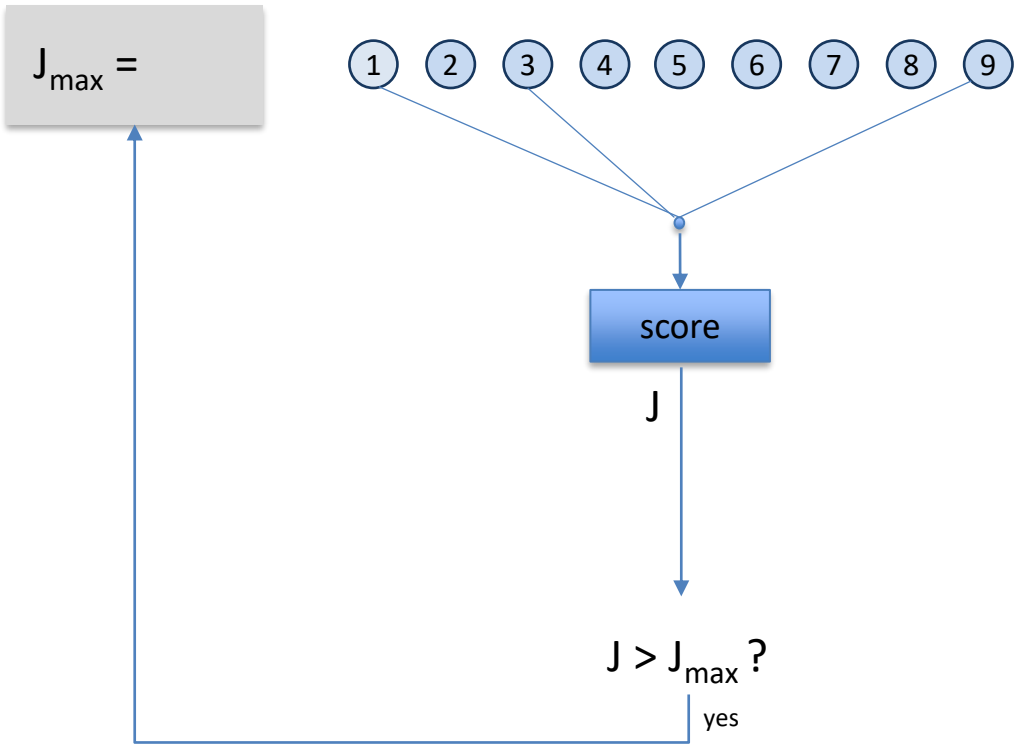

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

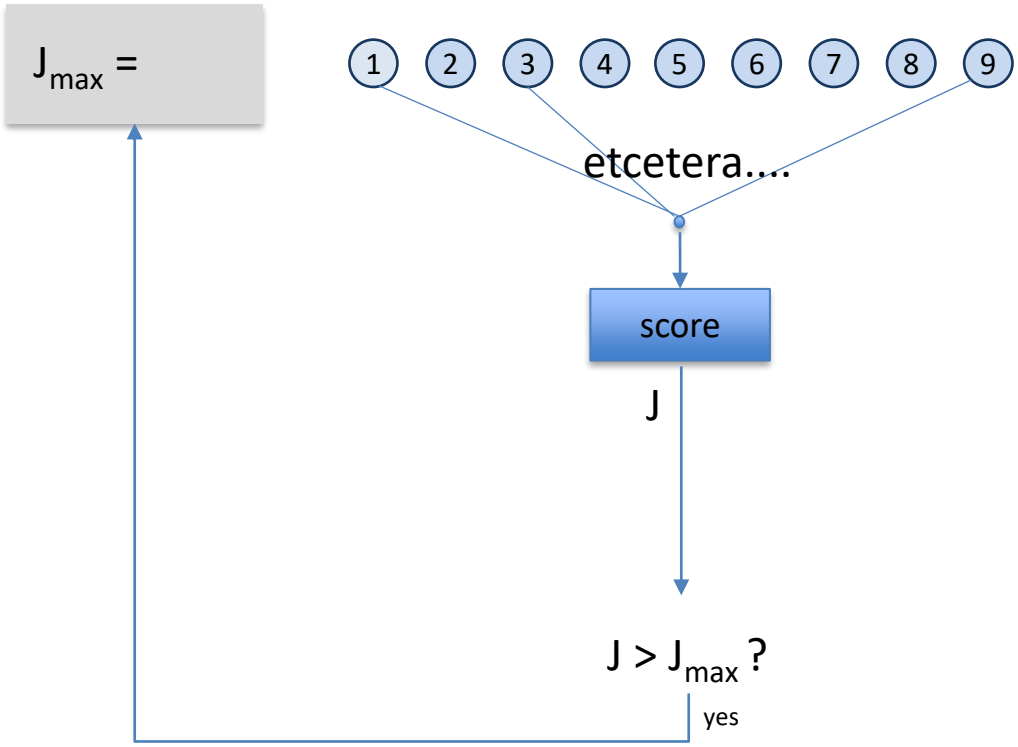

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

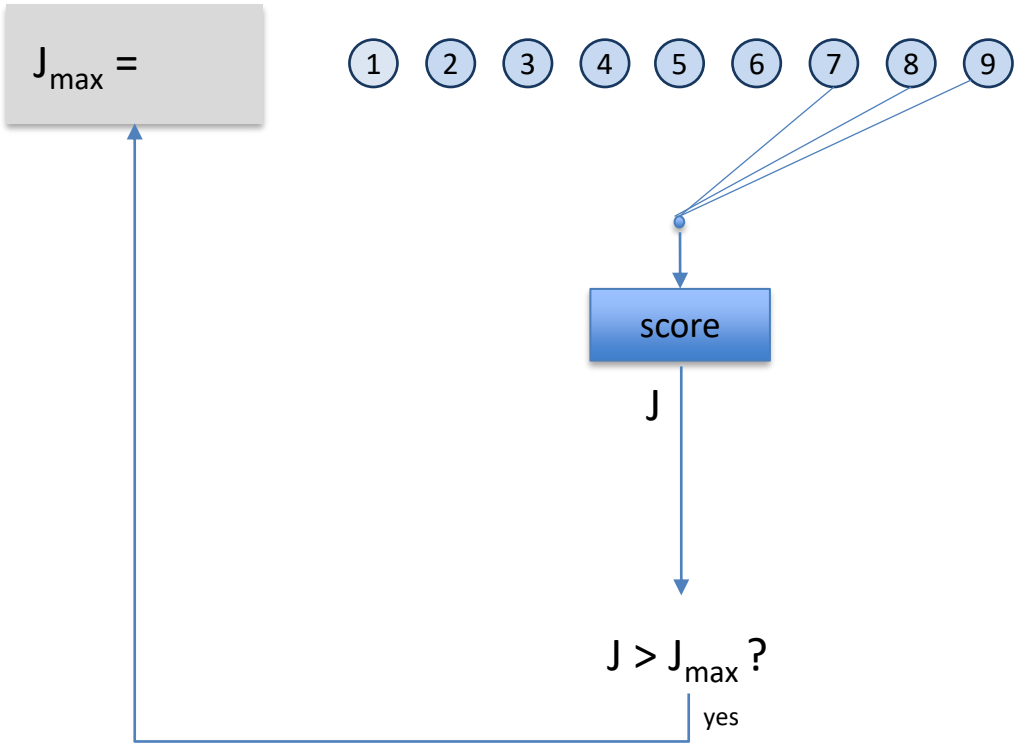

J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes


J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes


J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes


J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes


J_{max} =
1
2
3
4
5
6
7
8
9
etcetera....
score
J
$J > J_{max}$?
yes


J_{max} =
1
2
3
4
5
6
7
8
9
score
J
$J > J_{max}$?
yes

Exhaustive Search Algorithm - Example

Example: Selection of 3 features from 9

```
score_max = 0
for i1 = 1,2, … 7
    for i2 = i1+1, … 8
        for i3 = i2+1,… 9
            Xsel = X(:,(i1,i2,i3))
            score = fscore(Xsel)  # evaluation of the score
            if score>score_max
                sel_indices = (i1,i2,i3)
                score_max = score     # update maximum
            end
        end
    end
end

OUTPUT: sel_indices
```

Exhaustive Search Algorithm - General

```
Selection of p features from m

S   = all combinations of p numbers taken from m numbers*
N   = number of combinations
score_max = 0                  # initialization
for j = 1 to N               # evaluation of all combinations
    sj = S(j,:)              # columns of X
    score = fscore(X(:,st))   # evaluation of the score
    if score>score_max
        sel_indices = sj
        score_max = score     # update maximum
    end
end

OUTPUT: sel_indices


* In Matlab: nchoosek, in Python: scipy.misc.comb
```